LEVEL IV

C

Research Memorandum 74-2

# CLUSTER ANALYSIS OF OCCUPATIONAL DATA WITH FOCUS ON TASK RATHER THAN PEOPLE

AD A076722

D D C
RECEIVED
NOV 15 1979
E

DDC FILE COPY

This document has been approved for public release and sale; its distribution is unlimited.

U. S. Army

Research Institute for the Behavioral and Social Sciences

January 1974

79 11 13 312

# DISPOSITION FORM

For use of this form, see AR 340-15, the proponent agency is TAGCEN.

| REFERENCE OR OFFICE SYMBOL | SUBJECT |
|---|---|
| PERI-TP | Clearance and Transmittal of Reports to DTIC |

TO DDC-DAA-1
ATTN: Mr. Schrecengost

FROM ARI Rsch Pub Group

DATE 8 Nov 79
Ms Price/48913

CMT 1

1. The reports listed on Inclosure 1 are approved for public release with unlimited distribution (50 numbered ARI Research Memorandums, 74-1 thru 76-30).

2. These are among the previously unrecorded ARI reports which you identified to us 22 June 1979 as not in your retrieval system. The accompanying box contains at least one copy of each report for your retention and reproduction.

Helen S. Price

HELEN S. PRICE
Research Publications Group
Army Research Institute

1 incl
List of reports, 1974-76

x 4

| Accession For | |
|---|---|
| NTIS GRA&I | ✓ |
| DDC TAB | ☐ |
| Unannounced | ☐ |
| Justification | |
| By | |
| Distribution/ | |
| Availability Codes | |
| Dist | Avail and/or special |
| A | |

DA FORM 1 FEB 62 2496 REPLACES DD FORM 96, WHICH IS OBSOLETE.

~~Army Project~~ Number
(16) 2T061101A91B

Career Development 017

(14) ARI-RM-74-2

Research Memorandum 74-2

(9) Research memo,

(6) CLUSTER ANALYSIS OF OCCUPATIONAL DATA
WITH FOCUS ON TASK RATHER THAN PEOPLE,


(10) William J. Edison, [redacted] Robert F. Eastman,
Bertha H. Cory, Work Unit Leader

~~Submitted by:~~
Edmund F. Fuchs, Chief
Individual Training and Manpower Development Technical Area


(11) Janu~~ary 1~~974

(12) 10

Approved by:

E. Ralph Dusek, Director
Individual Training and Performance
Research Laboratory

J. E. Uhlaner, Technical Director
U. S. Army Research Institute for
the Behavioral and Social Sciences

---

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the Army Research Institute.

408 010 Gm

## CLUSTER ANALYSIS OF OCCUPATIONAL DATA WITH FOCUS ON TASK RATHER THAN PEOPLE

A "job" may be defined as a grouping of tasks performed by an individual to accomplish some purpose within an organization. Usually, the tasks which make up a job have a meaningful relationship with one another. That is, they might involve similar skills and requirements or they might be related by environmental factors, such as physical and temporal proximity. The job analyst is concerned with identifying and systematically recording the behaviors performed by job incumbents. From the collection and analysis of these data by the job analyst, inferences can be drawn and useful recommendations made regarding such matters as personnel selection policies, training programs, planning manning tables and force studies. One of the important needs of the job analyst for accomplishing these ends is a method of grouping tasks into meaningfully useful clusters. One such method is cluster analysis, i.e. a technique by which entities are formed into relatively homogeneous groups, based on similarity measures. The usual procedure in such analysis is to measure a number of attributes of the entities and by pairwise comparisons of the entities (and/or subclusters of the entities) form clusters based on the similarity of their respective attributes.[1]

When applied in occupational analysis these techniques can cluster <u>individuals</u> (entities) on the basis of the <u>tasks</u> (attributes) they perform. The results of this process are clusters of people who perform similar jobs.[2]

### OBVERSE CLUSTER ANALYSIS

Obverse cluster analysis is a modification of the usual clustering procedures so that clusters of <u>tasks</u> are constructed on the basis of individuals who perform them. The task measurements used are the same as in a traditional method of clustering. However, in obverse clustering a task is clustered with another task depending upon how many, or few, individuals perform both tasks.

### COMPARISON OF TWO METHODS

A simple comparison of the two methods is provided by analyzing the illustrative data in Table 1.

---

1/ Bailey, D. E., and Tyron, R. C. <u>Cluster Analysis</u>. McGraw-Hill, 1970.

2/ Bottenberg, R. A., and Christal, R. E. An iterative technique for clustering criteria which retains optimum predictive efficiency, <u>The Journal of Experimental Education</u>, <u>36</u>, (4), Summer 1968, 28-34.

Table 1

ILLUSTRATIVE EXAMPLE OF DATA ON TASKS AND INDIVIDUALS

| Individuals | Tasks 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| A | 1 | 1 | 1 | 0 | 0 |
| B | 0 | 0 | 0 | 1 | 1 |
| C | 1 | 1 | 1 | 0 | 1 |
| D | 0 | 0 | 0 | 1 | 1 |
| E | 1 | 0 | 1 | 0 | 0 |

Assume that five individuals (A, B, C, D, E) have been asked if they perform each of five tasks (1, 2, 3, 4, 5). In Table 1, the answers are recorded as: 1 = do perform the task, 0 = do not perform the task. Conceptually, clustering may proceed in stages from a first stage where each individual is a cluster of one to a final stage where the total group of individuals are clustered together with successive intermediate stages which are determined by the relative pairwise similarity of individuals (and/or clusters) to one another.

## CLUSTERING INDIVIDUALS

If one wants to cluster the individuals of the example above (a common clustering objective in personnel research), the similarity measure one would use is the number of tasks individuals perform in common. Thus, in the first stage individuals A and C would be clustered together ($C_1$), because they perform three tasks in common. In the next stage individual E would be added to $C_1$ because he is more similar to the members of this cluster than to either of the other individuals. Finally, individuals B and D would join together to form cluster $C_2$, because they are more similar to one another than to members of $C_1$. For job analysts some grouping between individuals and the total group may be useful for identifying individuals who perform similar jobs.

## OBVERSE CLUSTER ANALYSIS

To perform an obverse cluster analysis of the data in Table 1, the measure of similarity among tasks would be the number of individuals who perform a pair of tasks. For example, tasks 1 and 3 are both performed by three individuals (A, C, E), so they would form the first obverse cluster ($OC_1$) in the matrix. In the next stage task 2 would be added to $OC_1$ because it is performed by two individuals (A and C) who perform the member tasks in $OC_1$. Finally, tasks 4 and 5 would be clustered together to form a cluster ($OC_2$), because they are more similar to one another than to the tasks in $OC_1$.

This example illustrates that the same matrix can be employed to cluster either individuals or tasks. It should be noted that the matrix usually is rectangular, i.e., the number of individuals need not be the same as the number of tasks.

## THE CODAP SYSTEM

The Department of Defense is supporting a set of computer programs for occupational analysis called the Computerized Occupational Data Analysis Programs (CODAP) system. The CODAP system was developed by the Air Force[3] and is being used by Navy and Marine Corps. One part of the CODAP system is a clustering program to identify individuals or groups of individuals who perform similar jobs. Two measures of job similarity may be used as a basis for clustering, either the percentage of time that each individual devotes to each task he performs or the list of tasks he performs on the job.

## THE CODAP PROCEDURE FOR CLUSTERING INDIVIDUALS

To accomplish the clustering of individuals, the program casts the N individuals into an NxN table called a "similarity matrix" where the entries are the percentage of overlap in the tasks performed by each individual with each of the other individuals. The program then scans the similarity matrix and locates the overlap value of the two individuals whose task inventories indicate the greatest percentage of overlap in their jobs and combines them into a "cluster." The number of overlap values in the similarity matrix is thereby reduced to (N-1) x (N-1). In the next stage the program clusters the two most similar of the remaining N-1 entities (i.e., N-2 individuals and one cluster) in the matrix by using the criterion of the entry with the greatest percentage of overlap. This may involve either adding an individual to the existing cluster or uniting two individuals to form a new cluster. Either union will reduce the number of entities to N-2 and the similarity matrix will be collapsed to (N-2) x (N-2). The process continues in successive stages by combining the pair with the most similar sets of tasks performed (i.e., the greatest percentage of overlap in the similarity matrix) until the set of N individuals is expressed as a small group of clusters and the similarity matrix is reduced appropriately.

After the clustering process is completed, the job analyst is provided with a printout which indicates both the percentage of overlap between the two groups clustered at each stage in the process and the average percentage of overlap among all members within the newly formed cluster. From the latter measure the analyst is able to determine at what stage in the clustering process he wants to examine the groups which have been formed. That is, he will identify the clusters which have the degree of individual similarity among members useful for the analyst's purpose. The analyst can also examine a printout which assigns a number to every individual in such a way that all individuals clustered at any stage are listed together and may be found within a certain sequence range.

---

[3] Christal, R. E., and J. H., Jr. The MAXOF Clustering Model. In Proceedings: Conference on cluster analysis of multivariate data, New Orleans, December 1966. Washington: Office of Naval Research.

## ARI OBVERSE CLUSTERING, AN ADDITION TO THE CODAP SYSTEM

The Army Research Institute (ARI) has designed a system which uses the same input as the CODAP system with tasks in an NxN similarity matrix, but which locates the tasks which the greatest number of individuals perform in common and combines them into a cluster. The task similarity matrix is then reduced to (N-1) x (N-1). In the next stage the program clusters the two most similar of the remaining N-1 tasks and/or subclusters based on the number of individuals who perform them. Analogous to the clustering of individuals described previously, obverse clustering involves adding a task to the cluster or uniting two tasks to form a new cluster and thereby reducing the similarity matrix to (N-2) x (N-2) entries for N-2 tasks and/or clusters until the entire set of N tasks are included in a single cluster containing all tasks. In addition to clusterings tasks, the ARI addition to the CODAP system outputs an ordered list of tasks which reflect the content of the clusters and provides information about hierarchies of task clusters (i.e., subclusters within clusters, etc.).

## AN IMMEDIATE USE FOR OBVERSE CLUSTERING

An immediate application for obverse clustering was provided by the development of duty modules as part of a contract with the American Institutes for Research (AIR), on "A Taxonomic Base for Future Management Information and Decision System."[4] Duty modules are groups of tasks that tend to "go together" in meaningful ways and which satisfy certain operational requirements of utility. The obverse clustering of tasks performed by incumbents within MOS was used as a comparison with duty modules which have been independently developed from job descriptions and expert judgments.

Preliminary data from task inventories administered by the Army Office of Personnel Operations to incumbents of MOS 11 D (Armor Reconnaissance Specialist) and contained in the Military Occupations Data Bank (MODB) were used in a tryout of obverse clustering. Table 2 summarizes the correspondence between some of the task clusters empirically identified by CODAP obverse analysis for 11 D incumbents and "duty modules" for the same MOS developed by the expert judgment of an AIR team of recently retired Army personnel.

---

[4] Stephenson, R. W. A taxonomic base for future management information and decision systems: A common language for resource and requirement planning. ARI Technical Research Note 244, October 1972.

Table 2

PERCENT OF TASKS IN EMPIRICALLY IDENTIFIED CODAP CLUSTERS THAT FALL INTO DERIVED DUTY MODULES

| CODAP Obverse Cluster Number | Number of Tasks In Cluster | DUTY MODULES | | | | | |
|---|---|---|---|---|---|---|---|
| | | ADMINISTRATION A | | | | TRAINING B | |
| | | 1 | 2 | 3 | 4 | 1 | 2 |
| 1 | 6 | 50 | 33 | | | | |
| 2 | 9 | 44 | | 33 | | | 22 |
| 3 | 9 | 33 | | | 55 | | |
| 4 | 4 | | 50 | | | | |
| 5 | 6 | | | | | | 50 |
| 6 | 13 | | 45 | | | | 15 |
| 7 | 10 | | | | | 60 | 20 |

The rows in Table 2 correspond to the CODAP task clusters based on the frequency with which the tasks are assigned to the same people in the field. That is, insofar as the lists of tasks are sufficiently comprehensive to cover all duties, these clusters represent one concept of the "real world" of actual assignment practices as reported by job incumbents. The columns in Table 2 represent some of the duty modules derived by experienced judgment for the same MOS. The entries in Table 2 indicate what percentage of the tasks in each of the empirically identified (i.e., CODAP obverse method) clusters are also contained in duty modules A-1 through B-2. Examples of the tasks included in duty modules A-1 through B-2 are provided in Table 3.

Table 3

EXAMPLES OF THE TASKS INCLUDED IN DUTY MODULES A-1 THROUGH B-2

| Duty Module | Title | Task Example |
|---|---|---|
| A-1 | Performs general administration at company level headquarters | Prepare unit morning report |
| A-2 | Performs unit supervision and control of personnel | Schedule leaves and passes |
| A-3 | Establishes and operates a unit mail room | Receive and distribute personal mail |
| A-4 | Types, files and performs general clerical operations | Cut stencils and ditto masters |
| B-1 | Conducts unit and individual training | Prepare lesson plans and training aids |
| B-2 | Supervises and coordinates training in the unit | Evaluate personnel and recommend training |

If the percentage agreement in Table 2 is high, one may conclude that duty modules are comparable to the actual current assignment of duties to individuals in the field. This conclusion is justified because the probability of tasks being clustered together is based on the frequency with which they are performed by the same people. It should be pointed out that the percentages in Table 2 are strongly affected by the following conditions:

1. The tasklist for a given MOS may or may not adequately sample the tasks performed by an incumbent.

2. The tasklist for a given MOS may or may not include all the tasks which make up the duty modules for that MOS.

3. Duty assignments may exist which are inappropriate to a duty position designation within the MOS; this will tend to reduce the apparent fit of empirically derived clusters with duty modules (i.e., assignment practices are not necessarily perfect and may reflect a variety of contingencies).

4. The actual assignments reported in the questionnaires may not have included important missions that might be critical in combat but not encountered by most soldiers during the period covered by the questionnaires.

In summary, for the data available to date, it appears that rationally derived duty modules do overlap to a moderate degree with empirical task clusters derived by CODAP obverse method, even when conditions may serve to mitigate this correspondence.

## OTHER APPLICATIONS

Other possible applications of obverse clustering include identifying unit and individual performance criteria based on actual duties performed and forecasting equipment and maintenance needs based on measures of use by personnel.